Nine Month Report

David Matthews

Supervisors: Dr James Anderson and Dr Ben MacArthur

April 2, 2014

Graph Theory has been an established area of discrete mathematics since 1736 when Euler solved the famous question regarding the bridges of Königsberg. More recently, in the 1940s and 50s Erdős and Rényi laid the foundations of the theory of random graphs seeking to answer fundamental questions about the nature of "most" graphs. Random graphs have been studied for their own sake and have been used to model a diverse set of real-world networks from the world wide web to the metabolism of *E. coli*. Recently, the advent of the world wide web and the accompanying (relatively) cheap and powerful computational power has led to a reemergence of graph and random graph theory under the guise of "Network Science".

During the last nine months I have divided my attention between two projects which both pertain to random trees. The first project is entitled *the expected automorphism group of random recursive trees* and is an investigation into the relationship between a variation of the usual Fibonacci sequence and a family of randomly generated trees. We will refer to this project as *Project 1*. The second project is a continuation of my MSc dissertation which focused on a graph theoretical model of the cerebral vasculature. We will refer to this project as *Project 2*. Project 2 came about because of an interdisciplinary collaboration between Dr James Anderson in mathematics, Dr Roxana Carare who is an Alzheimer's disease specialist from the Faculty of Medicine and myself. This collaboration was facilitated by the Web Science doctoral training centre.

1 Introduction to Project 1

Given a family of algebraic structures, A, it is important to know when $a_1, a_2 \in A$ can be considered "the same" and to make precise the notion of breaking A into simpler pieces. Therefore we make three vital definitions in this section: random recursive trees (RRTs), RRT isomorphisms and induced subtrees.

A graph G is an ordered pair of disjoint sets (V, E) such that E is a subset of unordered pairs of elements of V. We say that $V(G)$ is the set of vertices or nodes of G and that $E(G)$ is the set of edges of G [1].

A *random recursive tree* (RRT), T, with vertices $V(T) = \{v_1, \dots, v_n\}$ is a labelled, rooted tree obtained by assigning a root vertex v_1 then adding $n-1$ vertices one by one such that each new vertex is joined by an edge to a randomly and uniformly chosen existing vertex. A random recursive q-ary tree is a labelled, rooted tree built in the same way as a random recursive tree except each new vertex is attached uniformly at random to an existing vertex that has outdegree less than q [2]. We say that RRTs and random recursive q-ary trees are *increasing* trees.

It is natural to consider increasing trees as increasing tree *processes* which are nested sequences of rooted, labelled trees

$$T_1 \subset T_2 \subset \cdots \subset T_n$$

Where each T_t has precisely t nodes (and $(t-1)$ edges). At time t vertex v is chosen from $V(T_{t-1})$ according to the appropriate attachment model and a new vertex v_t is attached to T_{t-1} via the edge $\{(v, v_t)\}$.

Example 1.1. For a RRT the appropriate attachment model is uniformly random random attachment. Since there are $(n-1)!$ possible sequences $(T_t)_1^n$, the RRT processes space can be made into a probability space by choosing each possible sequence equiprobably.

Let T_n be a random recursive tree with root v_1. For any $v \in V(T_n)$ such that $v \neq v_1$ let the set of vertices U_v be such that the shortest path from $u \in U_v$ to v does not contain v_1. We say that $B(v) = U_v \cap v_1$ is the *branch* containing v. We define $B(v_1) = V(T_n)$.

Given an increasing tree T_n on n vertices labelled by the function $\phi : V(T) \to \{1, 2, \dots, n\}$ and a vertex $v \in V(T)$ then we can consider $\tilde{T}_v$ which is the induced subtree with vertices, $v_i \in B(v)$ such that $\phi(v_i) \geq \phi(v)$.

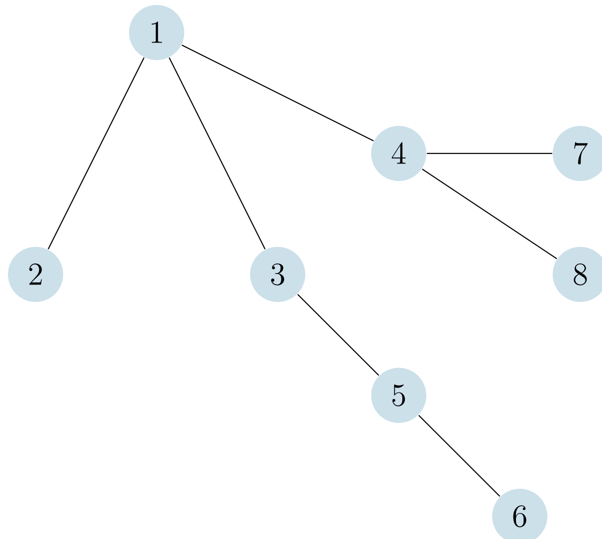

Figure 1: An example of a random recursive tree on 8 vertices. The induced subtree T_{v_4} consists of vertices v_4, v_5, v_6 and edges $\{(v_4, v_7), (v_4, v_8)\}$.

An automorphism of a graph is a permutation of the vertices of that graph which preserves adjacency (two vertices are said to be adjacent if there exists an edge joining them).

Numerical analysis conducted by MacArthur of 50 independent sequences in which the median value of $Aut(T_n)$ was calculated yielded the following conjecture.[1]

Conjecture 1. Let $(T_t)_0^n$ be a RRT process, then almost surely

$$\lim_{n\to\infty} |Aut(T_n)|^{\frac{1}{n}} \to \nu$$

where $\nu \sim 1.13198824$ is Viswanath's constant [3].

Viswanath's constant arises from the following simple adaptation of the Fibonacci sequence. Let $f_1 = f_2 = 1$ as usual and for $t > 2$ let $f_t = f_{t-2} \pm f_{t-1}$ where the plus and minus are chosen with probability 1/2. Viswanath showed that almost surely $|f(t)|^{\frac{1}{t}} \to 1.13198824$ as $t \to \infty$ although his proof was computational in nature [4]. We aim to prove Conjecture 1 and to derive Viswanath's constant analytically.

[1]We have confirmed MacArthur's numerical analysis using the graph isomorphism software *nauty*.

1.1 The automorphism group of random recursive trees

It is a result of Pólya that the automorphism group of a tree belongs to the class of permutation groups which contains the symmetric groups and is closed under taking direct and wreath products [5]. So there exists a direct product decomposition for any increasing tree T_n:

$$Aut(T_n) = A_1 \times A_2 \times \cdots \times A_p \times B_1 \times B_2 \times \cdots \times B_q \tag{1}$$

Such that $A_i \cong S_{x(i)}$ and $B_j = S_{y_1(j)} \wr S_{y_2(j)} \wr \cdots \wr S_{y_{k_j}(j)}$.

In addition to this algebraic interpretation of $Aut(T_n)$ there is a pleasing geometric realisation of equation 1 in which each symmetric factor A_i corresponds to a hub vertex adjacent to k paths of length n, each of which terminates in a leaf vertex which we say is a (n, k)-star. On the other hand each B_i corresponds to an extended symmetric induced subtree [6].

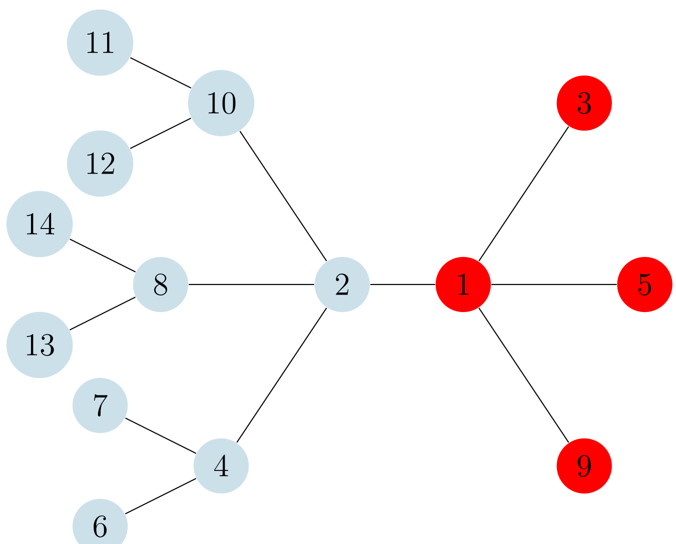

Figure 2: An example of a random recursive tree, T_{14} such that $Aut(T) \cong S_3 \times S_2 \wr S_3$. The red vertices indicate an induced subtree $\tilde{T}_{v_1}$ isomorphic to the bipartite graph $k_{1,3}$ which contributes S_3 to $Aut(T_{14})$. The blue nodes highlight an induced subtree, $\tilde{T}_{v_2}$ of T_{14} isomorphic to an extended symmetric subtree and contributes $S_2 \wr S_3$ to $Aut(T)$.

Example 1.2.

Let $(T_t)_1^n$ be a RRT process with root v_1 and labbeling ϕ. Consider any induced subtree $\tilde{T}_v \cong k_{1,p}$ of T_n with the vertex set $V(\tilde{T}_v) = \{v, v_1', v_2', \dots, v_p'\}$. By the definition of an induced subgraph $v_1', v_2', \dots, v_p'$ are leaves of T_n. If either $v_1 \notin V(\tilde{T}_v)$ or $v = v_1$ then $\phi(v) < \phi(v_j')$ for all j. However, if $v_1 \in V(\tilde{T}_v) \diagdown \{v\}$ then $\phi(v) \not< \phi(v_j')$ for all j.

For $t > 1$ the probability that v_1 is a leaf is $\frac{1}{t-1}$. So as $n \to \infty$:

$$\mathbb{P}(\text{ that there exists a } \tilde{T}_v \text{ such that } \phi(v) \not< \phi(v_j') \text{ for all } j \to 0) \to 1$$

1.2 Generalised Pólya Urns

The decomposition of automorphism group described by Equation 1 provides a way of proving Conjecture 1. It is enough to enumerate the expected proportion of (n, k)-stars and extended symmetric branches of an RRT to calculate the expected automorphism group. Since Pólya urn processes (defined below) have been used to calculate various features of RRTs such as degree distribution and expected depth of vertices they are a viable method for this enumeration.

A generalised Pólya urn process, $(X_t)_{t=0}^{\infty}$, is defined to have q types $1, 2, \dots, q$ and $X_t = (X_{t1}, \dots, X_{tq})$ such that X_{ti} is the number of balls of type i in the urn at time t. Initially the urn's contents are described by X_0 (deterministic or random). For each type i we associate an activity a_i and an expectation transition vector $\mathbb{E}(\xi_i) = (\xi_{i1}, \dots, \xi_{iq})$. At each time $t \geq 1$ one of the balls is chosen at random from the urn and the probability of drawing a ball of type i at time t is

$$\frac{a_i X_{(t-1),i}}{\sum_j a_j X_{(t-1),j}}.$$

Note that if every $a_i = 1$ then the balls are chosen uniformly at random. Once a ball of type i has been chosen it is returned to the urn with Y_{tj} balls of type j where $\mathbb{E}(Y_{tj}) = \xi_{ij}$ [7].

Let A denote the $q \times q$ matrix

$$A = \left(a_j \mathbb{E}(\xi_{ji})\right)_{i,j=1}^{q}.$$

Let Λ be the set of eigenvalues of A. The eigenspectrum of A will play a central role in the limiting behaviour of the urn process.

1.3 Properties of A

We write $i \succ j$ if $(A^n)_{ji} > 0$ for some $n \geq 0$. Note that $\succ$ is transitive and reflexive so it partitions the set of types into equivalence classes C_p such that i, j are in the same equivalence class if $i \prec j \succ i$. We say that type i is *dominating* if $i \succ j$ for every type j. Similarly a class C_k is said to be *dominating* if every $i \in C_k$ is dominating.

The usual assumptions one makes for an urn process are:

(A1) $\xi_{ij} + \delta_{ij} \geq 0$ a.s. for all i, j.

(A2) $\mathbb{E}(\xi_{ij}) < \infty$ for all i, j.

(A3) There exists a largest real eigenvector λ_1 such that $\lambda_1 > 0$.

(A4) λ_1 is simple.

(A5) There exists a dominating type, i and $X_{0i} > 0$.

(A6) λ_1 belongs to the dominating class.

An urn process which satisfies (A1)-(A6) is called irreducible and tenable. We say that the left and right eigenvectors corresponding to λ_1 are u_1 and v_1 respectively.

1.4 Limiting Behaviour

A generalised Pólya urn scheme can be embedded in a multi-type continuous time Markov branching process $\chi(t) = (\chi_1(t), \ldots, \chi_q(t))$ with initial vector $\chi(0) = X_0$. In this process particles of type i live for an expected time of a_i^{-1} (with exponential distribution). When a particle of type i dies it is replaced with a set of particles with distribution given by $(\xi_{ij} + \delta_{ij})_{j=1}^{q}$ [8].

Let $\tau_0 = 0$ and $\tau_n, n \geq 1$ be the n^{th} time a ball dies. The Pólya urn process $(X_n)_{n=0}^{\infty}$ equals in distribution the process $(\chi(\tau_n))_{n=0}^{\infty}$ so limit theorems for X_n can be derived from limit theorems for $\chi(t)$ - see Chapter V of [8].

Assume that A is tenable and irreducible. By Theorem 7.6 in [9], there exists a direct sum decomposition of $\mathbb{C}^q$ as $\bigoplus E_\lambda$ of eigenspaces E_λ and there exist projections P_λ for every eigenvalue λ of A that satisfy $\sum_{\lambda \in \Lambda} P_\lambda = I$ and $AP_\lambda = \lambda P_\lambda + N_\lambda$. Where N_λ is nilpotent. Let $d_\lambda \geq 0$ be the least integer such that $N_\lambda^{d_\lambda} \neq 0$. For $k = 0, 1, 2, \ldots$ we define the quotient space $E_{\lambda,k} := E_\lambda / N_\lambda^{k+1} E_\lambda$ and the projection $Q_{\lambda,k} : E_\lambda \to E_{\lambda,k}$.

In particular P_{λ_1} and $Q_{\lambda_1,0}$ are [7]:

$$P_{\lambda_1} = v_1 u_1^T,\ Q_{\lambda_1,0} = I \tag{2}$$

In [7], Lemma 9.2, Janson defines the following martingale if $t \geq 0$.

$$\mathcal{Y}(t) = e^{-tA}\chi(t) = \sum_{j=0}^{\infty} \frac{t^j A^j}{j!}\chi(t)$$

Lemma 1.3. *If $Re\lambda > \lambda_1/2$ for any $\lambda \in \Lambda$ then there exists a random vector $\tilde{W}_\lambda \in E_\lambda$ such that $P_\lambda \mathcal{Y}(t) \to \tilde{W}_\lambda$ almost surely as $t \to \infty$*

Proof. See Lemma 9.5 in [7] □

We also make the following definitions:

$$W_\lambda = Q_{\lambda,0}\tilde{W}_\lambda. \tag{3}$$

$$W = u_1 \cdot W_{\lambda_1} \tag{4}$$

Theorem 1.4. *[10] If a generalised Pólya urn is irreducible and tenable and $\sum a_j \xi_j = m$ for $j = 0, 1, \ldots, q$ and $m > 0$ then almost surely*

$$t^{-1}X_t \to \lambda_1 v_1 \text{ as } t \to \infty$$

Since the choice of v_1 is unique up to scalar factor we normalise such that $a \cdot v_1 = 1$.

In section 2 we will relax condition (A1) and just require that the urn process never allows balls to be removed from the urn unless those balls exist. In order that Theorem 1.4 holds we must show ([7], Remark 4.2):

(B1) There exists left and right eigenvectors u_1 and v_1 of λ_1 such that for every i in the dominating class $v_{1i} > 0$ and $u_{1i} > 0$.

(B2) $\mathbb{P}(u_1.\mathcal{Y}(t) > 0$ and $W = 0) \to 0$ almost surely as $t \to \infty$.

(B3) (A2)-(A6) hold.

2 $(1, k)$-stars

In this section we will describe a Pólya urn which will calculate the number of 1-stars of a random recursive q-ary tree, T_n^q for some fixed $q > 3$, rooted at v_0 with labelling ϕ. Balls in this urn scheme take exactly one of $q+1$ types and each ball corresponds to a node in a random recursive q-ary tree. Balls of types $j = 1, 2, 3, \ldots, q$ correspond to a vertex of an induced subtree $\tilde{T}_v \cong k_{1,j}$ with vertex set $V(\tilde{T}_v) = \{v, v_1', v_2', \ldots, v_j'\}$ such that $\phi(v) < \phi(v_p')$ for $p = 1, 2, \ldots, q$. Balls of type 0 correspond to any other node in $V(T_n)$.

Example 2.1. The Pólya urn corresponding to T_{14} depicted in Figure 2 would contain 1 ball of type 0, 9 balls of type 2 and 4 balls of type 3.

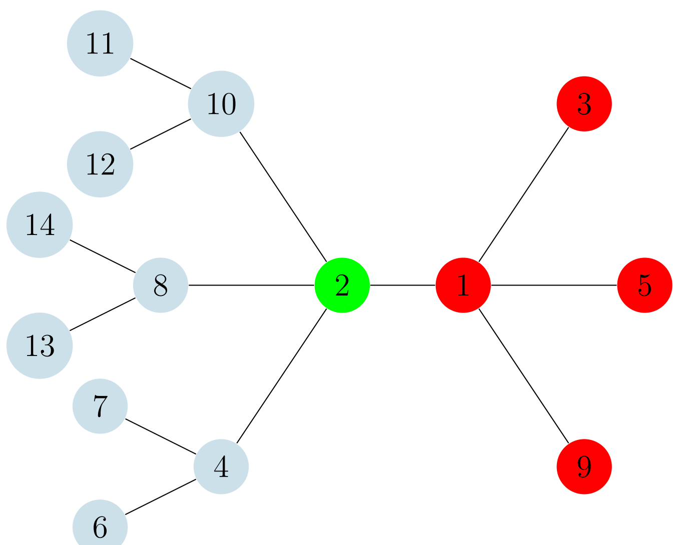

Figure 3: The 4 red vertices indicate an induced subtree $\tilde{T}_{v_1}$ isomorphic to the bipartite graph $k_{1,3}$ which contributes 4 balls of type 3 to the urn. The blue nodes indicate 3 induced subtrees of T_{14} isomorphic to $k_{1,3}$ and contribute 9 balls of type 2 to the urn. The green node corresponds to 1 ball of type 0 in the generalised Pólya urn.

At time t for $t = 0, 1, 2, 3, \ldots$ the contents of the urn are described by the $q+1$ dimensional vector $X_t = (X_{t0}, X_{t1}, \ldots X_{tq})$ such that each X_{tj} is defined to be the number of balls of type j in the urn at time t. Initially we set $X_0 = (1, 0, 0, \ldots, 0)$ which corresponds to the tree with one vertex and no edges.

We also assign an activity $a_j = 1$ to every type j (recall that this means that balls are picked from the urn uniformly at random) and an expectation transition vector $\mathbb{E}(\xi_j)$ to every type in the way we described in section 1.2. Each expectation transition vector is of the form $\mathbb{E}(\xi_j) = (\xi_{j0}, \xi_{j1}, \ldots, \xi_{jq})$, where $\xi_{jk}+\delta_{jk}$ corresponds to the expected number of nodes of type k created when a vertex of type j is chosen for attachment.

Note that the expectation transition vector has a general form for $j = 3, 4, \ldots, q-1$ and is more complicated otherwise. We will deal with the atypical cases first.

If a ball of type 1 is drawn from the urn at time t this corresponds to node v_t being attached to some induced subtree $\tilde{T}' \cong k_{1,1}$. Let $V(\tilde{T}') = \{v_1, v_2\}$ such that $\phi(v_1) < \phi(v_2)$, the probability of attaching v_t to v_1 is 0.5 and the probability of attaching to v_2 is 0.5. If v_t is attached to v_1 then the three vertices $\{v_1, v_2, v\}$ become part of a new subtree $\tilde{T}^{(2)} \cong k_{1,2}$. If v_t is attached to v_2 then v_1 corresponds to ball of type 0 and v_1 and v are the vertices of a new induced subtree $\tilde{T}^{(3)} \cong k_{1,1}$. This can be expressed in terms of ξ_{ij} as follows.

$$\mathbb{E}(\xi_1) = 1/2(1, 0, \ldots, 0) + 1/2(0, -2, 3, 0 \ldots, 0)$$

In a similar way we found that:

$$\mathbb{E}(\xi_0) = (-1, 2, 0, \ldots, 0) \tag{5}$$
$$\mathbb{E}(\xi_1) = (\frac{1}{2}, -1, \frac{3}{2}, 0, \ldots, 0) \tag{6}$$
$$\mathbb{E}(\xi_2) = (0, \frac{8}{3}, -3, \frac{4}{3}, 0, \ldots, 0) \tag{7}$$
$$\mathbb{E}(\xi_q) = (0, \frac{2q}{(q+1)}, 0, \ldots, 0, \frac{q^2}{(q+1)}, -q) \tag{8}$$
$$\tag{9}$$

On the other hand, if one was to pick a ball of type $j = 3, 4, \ldots q-1$ from the urn at time t this corresponds to a probability of $\frac{1}{j+1}$ that one attaches v_t to the hub of the corresponding induced subgraph $T' \cong k_{1,j}$ and a probability of $\frac{j}{j+1}$ that one attaches v_t to a leaf of T. Therefore, the appropriate transition vectors are of the form:

$$\xi_{j1} = \frac{2j}{j+1} \tag{10}$$
$$\xi_{j(j-1)} = \frac{j^2}{(j+1)}, \tag{11}$$
$$\xi_{jj} = -(j+1) \tag{12}$$
$$\xi_{j(j+1)} = \frac{(j+2)}{(j+1)} \tag{13}$$
$$\xi_{jk} = 0 \text{ otherwise} \tag{14}$$
$$\tag{15}$$

We can then make matrix A_q as described in section 1.2.

Example 2.2.

$$A_5 = \begin{pmatrix} -1 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 2 & -1 & \frac{8}{3} & \frac{6}{4} & \frac{8}{5} & \frac{10}{6} \\ 0 & \frac{3}{2} & -3 & \frac{9}{4} & 0 & 0 \\ 0 & 0 & \frac{4}{3} & -4 & \frac{16}{5} & 0 \\ 0 & 0 & 0 & \frac{5}{4} & -5 & \frac{25}{6} \\ 0 & 0 & 0 & 0 & \frac{6}{5} & -5 \end{pmatrix} \tag{16}$$

2.1 Bounds for v_1

To find v_1 we must solve $A_q v_1 = \mathbf{1} v_1$, which we can consider as a set of $q+1$ simultaneous equations which we will label $(E_0) - (E_q)$. We find bounds for the v_j using a series of inequalities.

Note that by E_0 and E_q:

$$v_1 = 4v_0 \text{ and } v_{q-1} = qv_q. \tag{17}$$

Lemma 2.3. *For $j = 1, 2, \ldots, q-2$*

$$v_j > \frac{q! v_q}{j!}$$

Proof. By E_{q-1}:

$$\frac{q}{q-1} v_{q-2} + \frac{q^2}{q+1} v_q = (q+1) v_{q-1}$$

By re-aranging and using equation 17 we found that,

$$\frac{v_{q-2}}{q-1} = (1 + \frac{1}{q} - \frac{1}{q+1})v_{q-1} > v_{q-1}. \tag{18}$$

We assume for the inductive hypothesis that for $n = 1, 2, 3, 4, \ldots, k-1$:

$$\frac{v_{q-n-2}}{q-n-1} > v_{q-n-1}$$

We have already proved the base case (equation 18) when $n = 0$.

Consider E_{q-k-1}:

$$\frac{q-k}{q-k-1}v_{q-k-2} + \frac{(q-k)^2}{q-k+1}v_{q-k} = (q-k+1)v_{q-k-1}$$

By the inductive hypothesis we have that:

$$\frac{v_{q-k-2}}{q-k-1} > (1 + \frac{1}{q-k} - \frac{1}{q-k+1})v_{q-k-1} > v_{q-k-1}$$

So for $j = 1, 2, 3, \ldots, q-2$

$$v_j > (j+1)v_{j+1} > \cdots > (j+1)(j+2)\ldots(q-1)v_{q-1} = \frac{q!}{j!}v_q \tag{19}$$

□

Lemma 2.4. *For* $j = 1, 2, 3, \ldots, q-2$

$$v_j < \frac{qq!(j+1)}{(j+1)!(q+1)}v_q$$

Proof. We will prove this lemma in a similar manner to lemma 2.3. For the base case we can rearange E_{q-1} and use equation 17 to find that $v_{q-2} < \frac{q^2}{q+1}v_{q-1}$.

For the inductive hypothesis assume that $v_{q-n-1} < \frac{(q-n+1)^2}{q-n+2}v_{q-n}$ for $n =$

$1, 2, 3, \ldots k$. Using the inductive hypothesis we can rearrange E_{q-k-1}:

$$\begin{aligned}
v_{q-k-2} &= \frac{q-k-1}{q-k}\left((q-k+1)v_{q-k-1} - \frac{(q-k)^2}{(q-k+1)}v_{q-k}\right) && (20)\\
&< \frac{q-k-1}{q-k}\left((q-k+1) - \frac{(q-k)^2(q-k+2)}{(q-k+1)^3}\right)v_{q-k-1} && (21)\\
&= \left(q-k+1-1-\frac{1}{(q-k)} - \frac{(q-k)^2(q-k+2)}{(q-k+1)^3}\right)v_{q-k-1} && (22)\\
&= \left(\frac{(q-k+1)((q-k)^2-1)}{(q-k)(q-k+1)} - \frac{(q-k)^2(q-k+2)}{(q-k+1)^3}\right)v_{q-k-1} && (23)\\
&= \left(\frac{(q-k)^2}{q-k+1} + \frac{1}{(q-k)(q-k+1)} - \frac{1}{q-k+1} - \frac{(q-k)^2(q-k+2)}{(q-k+1)^3}\right)v_{q-k-1} && (24)\\
&< \frac{(q-k)^2}{q-k+1}v_{q-k-1} && (25)
\end{aligned}$$

Therefore for any $j = 1, 2, \ldots, q-2$, by equation 20,

$$\begin{aligned}
v_j &< \frac{(j+2)^2}{(j+3)}v_{j+1} && (26)\\
&< \frac{(j+2)^2(j+3)}{(j+4)}v_{j+2} < \cdots < \frac{(j+2)^2(j+3)\ldots(q-1)q}{(q+1)}v_{q-1} && (27)\\
&= \frac{(j+2)^2(j+3)\ldots(q-1)q^2}{(q+1)}v_q && (28)\\
v_j &< \frac{q!q(j+2)}{(j+1)!(q+1)}v_q. && (29)
\end{aligned}$$

□

Lemma 2.5. *For $q > X$ the following inequality holds:*

$$\frac{8}{23q!} < \frac{8(q+1)}{23q!q} < v_q < \frac{4}{7q!}$$

Proof. By the normalisation property discussed in Theorem 1.4 $\sum_{j=0}^{q} v_i = 1$.

Using Equation 17 with Lemmas 2.3 and 2.4 we found that:

$$1 = \sum_{j=0}^{q} v_j < \sum_{j=1}^{q-2} \frac{qq!(j+2)v_q}{(j+1)!(q+1)} + v_0 + v_{q-1} + v_q \tag{30}$$

$$< \left(\frac{qq!}{q+1} \sum_{j=1}^{q-2} \frac{(j+2)}{j+1)!} + \frac{3q!q}{8(q+1)} + q + 1 \right) v_q \tag{31}$$

$$< \left(\frac{qq!}{q+1}(1 + 2(e-2)) + \frac{3q!q}{8(q+1)} + q + 1 \right) v_q \tag{32}$$

$$< \frac{23q!q}{8(q+1)} \tag{33}$$

$$1 = \sum_{j=0}^{q} v_j > \sum_{j=1}^{q-2} \frac{q!v_q}{j!} + v_0 + v_{q-1} + v_q \tag{34}$$

$$> q!v_q \sum_{j=1}^{q-2} \frac{1}{j!} + \frac{q!}{4} v_q + qv_q + 1v_q \tag{35}$$

$$> \frac{7}{4} qq!v_q \tag{36}$$

A simple rearangement completes the proof of Lemma 2.5. □

If we collect results from Lemmas 2.3, 2.4 and 2.5 we see that for $j = 1, 2, \ldots j-2$:

$$\frac{8}{23j!} < v_j < \frac{4q(j+1)}{7(j+1)!(q+1)} < \frac{4}{7j!} \tag{37}$$

Lemma 2.6. *(B1)-(B3) hold for A.*

Proof. Since $T = A + \alpha I$ is non-negative if $\alpha > q$, by Perron -Frobenius theory T is a nonnegative, irreducible matrix. Therefore A has a real, simple eigenvalue λ_1 (A4) such that for any other eigenvalue λ, $Re(\lambda) < \lambda_1$ and the corresponding left and right eigenvectors are the only positive eigenvectors of A [11] . By inspection $u_1 = (1, 1, \ldots, 1)$ is a left eigenvector of A corresponding to eigenvalue 1 (A3), since $u_{1i} > 0$ for all i, $\lambda_1 = 1$. The bounds calculated above show that there exists a right eigenvector v_1 corrosponding to λ_1 such that $v_{1j} > 0$ for $j = 0, 1, 2, \ldots q$.

By the defiition of this urn process (A2) is true. Every type is dominating so, in particular, type 0 is dominating and $X_{00} = 1 > 0$ (A5). There is one dominating class so λ_1 belongs to the dominating class (A6). Therefore (B1) and (B3) are satisfied.

For (B2) assume that $0 = W = u_1 W_{\lambda_1} = \mathbf{1} W_{\lambda_1} = \mathbf{1} Q_{\lambda_1,0} \tilde{W}_{\lambda_1} \mathbf{1} I \tilde{W}_{\lambda_1}$ so $0 = \tilde{W}_{\lambda_1}$.

Note that $Re(\lambda_1) = 1 > \frac{1}{2} = \frac{\lambda_1}{2}$ so, by Lemma 1.3 and equation 2 almost surely $lim_{t\to\infty} P_{\lambda_1} y(t) \to 0$ we can rewrite this as almost surely

$$lim_{t\to\infty}(v_1 u_1^T) y(t) \to 0$$

Therefore almost surely,

$$lim_{t\to\infty} y(t) \to 0.$$

This means that almost surely $\mathbb{P}(u_1 \dot{y}(t) > 0 \text{ and } W = 0) \to 0$ as $t \to \infty$. $\square$

Also note that $\sum_{j=0}^{q} a_j \xi_j = 1$ which means that one ball is added each time (corresponding to making one attachment at each discrete time). Therefore we can appeal to Theorem 1.4 to see that:

$$t^{-1} X_t \to \lambda_1 v_1 \text{ almost surely, as } t \to \infty$$

2.2 Automorphisms coming from 1-stars

To achieve our glorious goal; the calculation of the cardinality of the expected automorphism of 'most' RRTs our method has been to enumerate the expected number of (n, k)-stars and extended symmetric branches. As a preliminary stage we consider the subgroup, Aut_1 of the automorphism group of a random recursive $q-ary$ process on n nodes that corrosponds to $(1, k)$-stars.

Theorem 2.7. *Most random recursive q-ary tree processes T_n^q, are such that $lim_{\min(n-1,q)\to\infty} |Aut_1(T_n^q)|^{\frac{1}{n}} < \infty$.*

Proof. In section 2.1 we found bounds for the primary eigenvector, $v_1 = (v_{10}, \dots v_{1q})$, of the matrix associated with a Pólya urn process that corresponds with a random recursive q-ary tree process. By Lemma 2.6 v_{1j} is the expected proportion of nodes that are contained in induced subtrees $\tilde{T}_v \cong k_{1,j}$ of T_n^q such that those vertices respect the ordering decribed at the beginning of section 2 in the limit as $n \to \infty$. As n grows this ordering no longer becomes a restriction in the limit as $n \to \infty$ the number of expected $(1, j)$-stars is $\frac{n v_{1j}}{(j+1)}$.

The contribution of a $(1,k)$-star to $Aut(T_n^q)$ is S_j so $Aut_1(T_n^q)$ can be written as the direct product of symmetric groups. Further $|Aut_1(T_n^q)|$ can be written as the product of factorials. The extended Pólya urn process tells us that for most random recursive q-ary tree processes T_n^q,

$$lim_{n\to\infty}\mathbb{E}(|Aut_1(T_n^q)|) = \prod_{j=2}^{\min(n-1,q)} j!^{\frac{nv_j}{(j+1)}}$$

By Lemma 2.5 we immediately have the following bounds:

$$\left(\prod_{j=2}^{\min(n-1,q)} j!^{\frac{nv_j}{(j+1)}}\right)^{\frac{1}{n}} < \left(\prod_{j=2}^{\min(n-1,q)} j!^{\frac{4n}{(7(j+1)!}}\right)^{\frac{1}{n}} \tag{38}$$

$$\left(\prod_{j=2}^{\min(n-1,q)} j!^{\frac{nv_j}{(j+1)}}\right)^{\frac{1}{n}} < \left(\prod_{j=2}^{\min(n-1,q)} j!^{\frac{8n}{23(j+1)!}}\right)^{\frac{1}{n}} \tag{39}$$

Consider the series $S = \sum_{j=2}^{\infty} \ln\left(j!^{\frac{4}{7(j+1)!}}\right) = \frac{4}{7}\sum_{j=2}^{\infty} \frac{ln(j!)}{(j+1)!}$. Since S converges, by [12] it is true that $\prod_{j=2}^{\infty} j!^{\frac{4n}{7(j+1)!}} < \infty$.

□

To perform a reality check with this result please note that:

$$\prod_{i=2}^{q} i!^{\frac{1}{(i+1)!}} = 2^{\frac{1}{6}+\frac{1}{24}+\cdots}3^{\frac{1}{24}+\frac{1}{120}+\cdots}\cdots \sim 1.25. \tag{40}$$

Furthermore $1.25^{\frac{8}{23}} = 1.0807... < \nu = 1.13198824$ where ν is Viswanath's constant.

3 Enumerating $(2,k)$ -stars

In section 2 we described a Pólya urn process in order to calculate in the limit the expected number of $(1,k)$-stars of an RRT. In this section we will extend this result and use a Pólya urn method to calculate the expected number of $(2,k)$- stars.

Let T_n^q be a RRT for some fixed $q > 3$, rooted at v_0 with labelling ϕ. Balls in this urn scheme take exactly one of $\frac{1}{2}(q+1)(q+2)$ types and each ball corresponds to a node in a random recursive q-ary tree.

Balls of types $j = 1, 2, 3, \ldots, q$ correspond to a vertex of an induced subtree $\tilde{T}_v \cong k_{1,j}$ with vertex set $V(\tilde{T}_v) = \{v, v_1', v_2', \ldots, v_j'\}$ such that $\phi(v) < \phi(v_p')$ for $p = 1, 2, \ldots, q$.

Balls of type $k > 1$ can be written uniquely as

$$k = \sum_{j=1}^{x} (q + 1 - j) + y$$

where $0 < y < q + 1 - j$ correspond to vertices of induced subtrees of T with a hub-node v of outdegree $x + y$ such that v is adjacent to x paths of length two terminating in a leaf vertex and v is also adjacent to y leaves. Balls of type 0 correspond to any other node in $V(T_n)$.

Example 3.1. (i) A ball of type $k = 3 = 0 + 3$ corresponds to an induced subtree isomorphic to $k_{1,3}$.

(ii) A ball of type $k = q + 2 = \sum_{j=0}^{1}(q + 1 - j) + 1$ corresponds to an induced subtree isomorphic to the graph shown in figure 4.

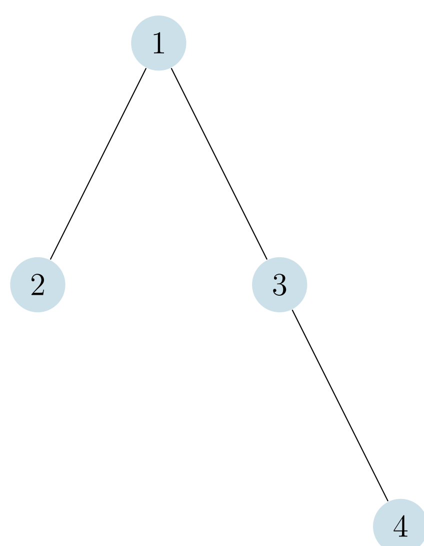

Figure 4

In the same way as in Section 2 we define the expected transition probability vectors xi_j for $j = 1, \ldots, \frac{1}{2}(q+1)(q+2)$.

The first thing to note is that $\xi_1 = (-1, 2, 0, \ldots, 0)$ again since if we attach to a vertex that corresponds to a ball of type 0 we get two balls of type 1. However $\xi_1, \xi_2, \ldots \xi_q$ have a different form to the transition vectors fro the Pólya urn described in section 2. Assume that at time t a vertex v is chosen that is contained in an induced subtree isomorphic to a $(1,k) - star$ where $0 < k < q$. Then there is a probability of $\frac{1}{k+1}$ that v is the hub of the $(1,k)$ - star in which case we form a $(1, k+1)$ - star i.e we add. There is a probability of $\frac{k}{k+1}$ that v is a leaf of the $(1,k)$ -star, which means that a new induced subtree with a hub adjacent to $(k-1)$ leaves and 1 path of length 2.

In terms of the transition vectors, for $0 < k < q$ the following are the only non-zero components of ξ_k:

$$\begin{aligned}
\xi_{kk} &= -(k+1) \\
\xi_{k(k+1)} &= \frac{k+2}{k+1} \\
\xi_{k(q+k)} &= \frac{k(k+2)}{k+1}
\end{aligned}$$

If, we choose a vertex corresponding to type $q+1$ then with probability $\frac{1}{3}$ we make a new induced subtree $\tilde{T}_v^y$ such that v is adjacent to a paths of length two terminating in a leaf vertex and a leaf. With probability $\frac{1}{3}$ we make a new induced subtree isomorphic to a $(k,2)$-star and we make one vertex with corresponding type 0. With probability $\frac{1}{3}$ we make a new induced subtree also corresponding to type $q+1$ and we make one vertex with type 0.

$$\xi_{q+1} = \left(\frac{2}{3}, 0, 1, 0, \ldots, 0, -2, \frac{4}{3}, 0, \ldots, 0 \right)$$

Assume that at time t we choose a vertex contained in an induced subtree, $\tilde{T}_v$, such that v is adjacent to x paths of length two terminating in a leaf vertex and 0 leaves such that $x < q$ and $x > 1$.

(i) With probability $\frac{1}{2x+1}$ we choose v and make a new induced subtree $\tilde{T}_v^v$ such that v is adjacent to x paths of length two terminating in a leaf vertex and 1 leaf vertex.

(ii) With probability $\frac{x}{2x+1}$ we choose a vertex distance 1 from v and make a new induced subtree $\tilde{T}_v^{x1}$ such that v is adjacent to $x-1$ paths of length two terminating in a leaf vertex and 1 leaf vertex and an induced subtree isomorphic to a $(1,2)-star$.

(iii) With probability $\frac{x}{2x+1}$ we choose a vertex distance 1 from v and make a new induced subtree $\tilde{T}_v^{x2}$ such that v is adjacent to $x-1$ paths of length two terminating in a leaf vertex and 1 leaf vertex and an induced subtree of type $q+1$.

If $k=\sum_{j=1}^{x}(q+1-j)+y$ such that $x>1$, $y=0$ thfacebooken the only non-zero components of ξ_k are as follows:

$$\begin{aligned}
\xi_{k2} &= \frac{3x}{2x+1} \\
\xi_{k(q+1)} &= \frac{3x}{2x+1} \\
\xi_{kj} &= \frac{2x(2x-1)}{(2x+1)} \\
\xi_{kk} &= -2x+1 \\
\xi_{k(k+1)} &= \frac{2x+2}{2x+1}
\end{aligned}$$

Such that $j=\sum_{j=1}^{x-1}(q+1-j)$.

Now assume that at time t we choose a vertex contained in an induced subtree, $\tilde{T}_v$, such that v is adjacent to x paths of length two terminating in a leaf vertex and y leaves such that $x+y<q$ and $x,y>0$.

(i) With probability $\frac{y}{2x+y+1}$ we choose a leaf vertex distance 1 from v and we make a new induced subtree $\tilde{T}_v^{y}$ such that v is adjacent to $x+1$ paths of length two terminating in a leaf vertex and $y-1$ leaves.

(ii) With probability $\frac{x}{2x+y+1}$ we choose a non-leaf vertex distance 1 from v and we make a new induced subtree $\tilde{T}_v^{x1}$ such that v is adjacent to $x-1$ paths of length two terminating in a leaf vertex and y leaves and we make an induced subtree isomorphic to a $(1,2)$-star.

(iii) With probability $\frac{x}{2x+y+1}$ we choose a vertex distance 2 from v and we make a new induced subtree $\tilde{T}_v^{x2}$ such that v is adjacent to $x-1$ paths of length two terminating in a leaf vertex and y leaves and we make an induced subtree isomorphic to a path of length 2.

(iv) With probability $\frac{1}{2x+y+1}$ we choose v and we make a new induced subtree $\tilde{T}_v^v$ such that v is adjacent to x paths of length two terminating in a leaf vertex and $y+1$ leaves.

Therefore if $k = \sum_{j=1}^{x}(q+1-j)+y$ such that $x, y > 0$, then the only non-zero components of ξ_k are as follows:

$$\begin{aligned}
\xi_{k2} &= \frac{3x}{2x+y+1} \\
\xi_{k(q+1)} &= \frac{3x}{2x+y+1} \\
\xi_{kj} &= \frac{2x(2x+y-1)}{2x+y+1} \\
\xi_{kk} &= -(2x+y+1) \\
\xi_{kk+1} &= \frac{2x+y+2}{2x+y+1} \\
\xi_{kl} &= \frac{y(2x+y+2)}{2x+y+1}
\end{aligned}$$

Such that $j = \sum_{j=1}^{x-1}(q+1-j)+y$ and $l = \sum_{j=1}^{x+1}(q+1-j)+y-1$.

For the final cases assume that at time t we choose a vertex of type k contained in an induced subtree, $\tilde{T}_v$, such that v is adjacent to x paths of length two terminating in a leaf vertex and y leaves such that $x+y=q$. Using a similar method to section 2 the non-zero components of ξ_k to be:

$$\begin{aligned}
\xi_{k2} &= \frac{3x}{2x+y+1} \\
\xi_{k(q+1)} &= \frac{3x}{2x+y+1} \\
\xi_{kj} &= \frac{2x(2x+y-1)}{2x+y+1} \\
\xi_{kk} &= -\frac{2x+y}{(2x+y+1)} \\
\xi_{kl} &= \frac{y(2x+y+2)}{2x+y+1}
\end{aligned}$$

Such that $j = \sum_{j=1}^{x-1}(q+1-j)+y$ and $l = \sum_{j=1}^{x+1}(q+1-j)+y-1$.

For simplicity we will now refer to type $k = \sum_{j=1}^{x}(q+1-j) + y$ as type $\tau_{x,y}$.

Following the method of section 2 this new Pólya urn is tenable and irreducible with principle eigenvalue $\lambda_1 = 1$. So consider the matrix A' associated with this process and let the principle right eigenvector be w. Note that $w_k = v_k$ for $k = 1, 2, \dots q$ so we have bounds for these values. Clearly $w_0 < v_0$.

4 Bounds for w

For simplicity we will now refer to type $k = \sum_{j=1}^{x}(q+1-j) + y$ as type $\tau_{x,y}$ we will further write $w_{x,y} := w_{\tau_{x,y}}$.

Each $w_{x,y}$ is the expected proportion of vertices that are contained in $V(\tilde{T}_v)$ where $\tilde{T}_v$ is an induced subtree of the random recursive q-ary tree T_n^q such that v is adjacent to x paths of length 2 and y leaves in the limit as $n \to \infty$. Therefore the expected limiting proportion of induced subtrees containing vertices of type $\tau_{x,y}$ is $w_{x,y}/(2x + y + 1)$. Since each induced subtree of type $\tau_{k,y}$ contributes to proportion of $(2, k)$-stars we can write this proportion, P, as:

$$P = \sum_{y=0}^{q-k} \frac{w_{k,y}}{2k + y + 1}$$

We can replicate the method from section 2.1and consider the eigen equation $A'w = 1w$ as a collection of $\frac{1}{2}(q+1)(q+2)$ simultaneous equations which we wil label E_k for $k = 0, 1, 2, \dots, \frac{1}{2}(q+1)(q+2) - 1$, which can be written as $E_{x,y}$ for simplicity. Our guiding principle will be to use the bounds for v that we formulated in section 2.1 to calculate bounds for w.

In particular, if $2 < k < q$ then the general form of E_k is as follows:

$$\frac{k+1}{k} w_{0,k-1} - (k+1)v_{0,k} + \frac{2(k+1)}{(k+3)} w_{1,k} = w_{0,k} \tag{41}$$

Since $w_{k,0} = v_k$ for $k = 1, 2, \dots q$ we can write equation 41 as:

$$\frac{k+1}{k} w_{0,k-1} - (k+1)v_{0,k} + \frac{2(k+1)}{(k+3)} w_{1,k} = w_{0,k}$$

If we collect terms and rearrange for $w(1, k)$ we find that:

$$w_{1,k} = \frac{(k+2)(k+3)}{2(k+1)} v_k - \frac{(k+3)}{2k} v_{k-1} \tag{42}$$

Using the bounds for v_{k-1} (equation 37) we find the following bounds for $w_{1,k}$.

$$w_{1,k} < \frac{(k+2)(k+3)}{2(k+1)} v_k - \frac{4(k+3)}{23k!} \tag{43}$$

We can also use the bounds for v_{k-1} (equation 37) to find the following improved bounds for $w_{1,k}$:

$$w_{1,k} < \frac{2(k+2)(k+3)}{7(k+1)!} - \frac{4(k+3)}{23k!} \tag{44}$$

If we rearrange and simplify equation 44 we find that:

$$w_{1,k} < \frac{(k+3)(18k-64)}{161(k+1)!}$$

References

[1] B. Bollobas, *Modern Graph Theory*, Corrected ed. (Springer, 1998).

[2] F. Bergeron, P. Flajolet, and B. Salvy, Varieties of increasing trees, in *CAAP '92*, edited by J.-C. Raoult, , Lecture Notes in Computer Science Vol. 581, pp. 24–48, Springer Berlin Heidelberg, 1992.

[3] B. MacArthur.

[4] D. Viswanath, Mathematics of Computation of the American Mathematical Society **69**, 1131 (2000).

[5] N. Biggs, *Algebraic Graph Theory*, 2nd ed. (Cambridge University Press, 1993).

[6] B. D. Macarthur, R. J. Sánchez-garcía, and J. W. Anderson, Discrete Applied Mathematics , 3525 (2008).

[7] S. Janson, APPL **110**, 177 (2004).

[8] K. Athreya and P. Ney, *Branching processes* (Dover Publications, 2004).

[9] K.Nomizu, *Fundamentals of linear algebra*, 2nd ed. (Chelsea Publishing Co., 1979).

[10] S. Janson, Random Structures and Algorithms **26**, 69 (2005).

[11] E. Seneta, *Non-negative Matrices and Markov Chains (Springer Series in Statistics)*, 2nd ed. (Springer, 2006).

[12] G. Jones and Singerman, *Complex Functions* (Cambridge University Press, Cambridge, UK, 1987).